

# **Anomaly detection in online social networks**

Savage, D., Zhang, X., Yu, X., Chou, P., Wang, Q., 2014

# Introduction



- Anomalies in social networks are often representative of **irregular** or **unwanted** behaviour
- Detection of anomalies has been used to identify:
  - Malicious individuals such as spammers, sexual predators, and online fraudsters
  - Important and influential nodes
- Process of detecting anomalies:
  - Calculation of a **suitable feature space**
  - Detection of **anomalies within this space**

# Anomalies



- **Patterns of interaction between individuals** of a network that significantly differ from the normal behaviour
  - Determine which features best differentiate normal and anomalous behaviour
- Characteristics of anomalies:
  - Dynamic or Static
  - Labelled or Unlabelled
  - Global or Local
    - Minimal anomalous unit

# Static anomalies



- **Static Unlabelled Anomaly**
  - When the behaviour of an individual or individuals leads to the formation of unusual network structures like a star or clique
- **Static Labelled Anomaly**
  - Provides a context of interaction
  - Typical network structures within a certain context may be considered malicious

# Dynamic anomalies



- **Dynamic Unlabelled Anomaly**
  - When patterns of interaction change over time
    - the structure of a network in one time-step differs from that in previous time-steps
- **Dynamic Labelled Anomaly**
  - Extends the techniques used in dynamic unlabelled anomalies with the additional information provided by the labels

# Indicators of suspicious activity



- **Star**
  - When a single individual is connected to nodes who were not connected otherwise
- **Number of triangles in ego-net**
  - Low triangle count possibly indicates anomalous behaviour
- **'Heavy' ego-net**
  - When the sum over a particular label is disproportionately high relative to the number of edges

Ego-net: subgraph of a subject and its immediate neighbours

# Process of detecting anomalies



1. Determine the **smallest unit affected** by the behaviour of interest
2. Identify the **particular properties** of this unit that are expected to deviate from the norm
3. Identify the **context** in which these deviations are expected
4. Calculate the properties of interest, **extracting the feature space**
5. Within this space, calculate **distances between observations**

# Problem selecting feature space



- The selection of a suitable feature space encompasses the points **1-4** of the previous slide
- Determining **which features will provide the greatest separation of normal and anomalous behaviour** is in itself a key challenge in anomaly detection
  - The **behaviour of anomalous entities may change over time** in direct response to the detection methods employed



# Conclusions



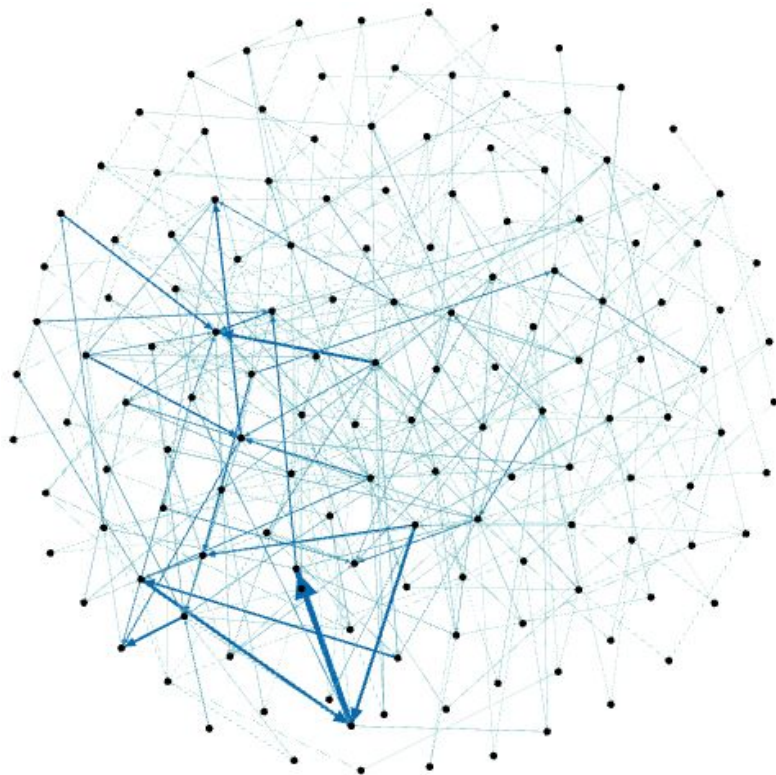
- Relatively **new field** that requires further research
- **Requirements** for anomaly detection in social networks will rapidly advance in the future
  - Larger volumes of data
  - More complex behaviours
- The **hardest task** is the **selection of a suitable feature space**

# Regarding the Football Transfer Market



- The football transfers market moves **large amounts of money** and so it is likely to lead clubs to perform somehow **strange operations**
- The network is static so the type of anomalies that fall in this context are the **static anomalies**
  - If we consider network snapshots from different seasons we can also consider the **dynamic anomalies**

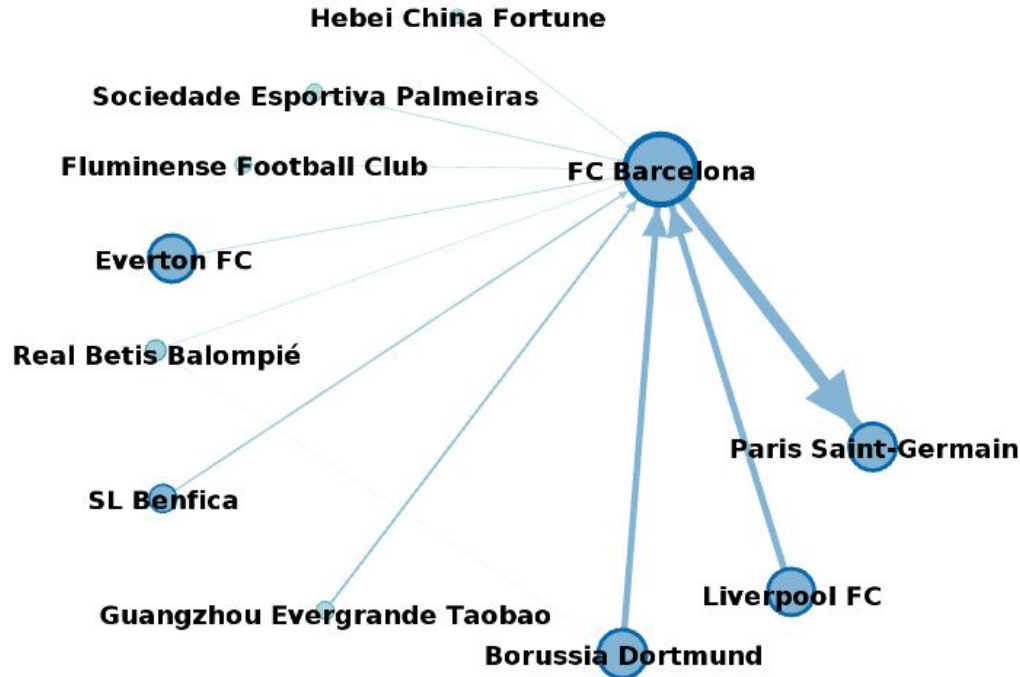
# 2017 Transfers



Avg. Weighted Degree 33178750

- Given the average weighted degree, the boldest transfer (222M) is out of the normal interaction patterns
  - It is about **6,7 times greater** than the average money moved by each club.
  - And it is just one transfer

# 2017 FC Barcelona Heavy Ego-Net



- FC Barcelona moved a total amount of **566M** euros between his **10 neighbours**
  - But moved **457M** euros between only **3** of his neighbours
  - This 3 transfers represent roughly **81%** of the total money

# 2015-2017 Leeds United Star Network



- Leeds United has made 9 transfers in 3 years
  - All of them with clubs that are **not linked together**
  - It has **0 triangles** in its ego-network
- This is not necessarily an anomaly, but it is a strange pattern of interaction